(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number
WO 01/55437 A2

(51) International Patent Classification[7]: C12P 19/34

(21) International Application Number: PCT/US01/02623

(22) International Filing Date: 25 January 2001 (25.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/491,404 25 January 2000 (25.01.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/491,404 (CIP)
Filed on 25 January 2000 (25.01.2000)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvalle, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US). **DRMANAC, Rodje, T.** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/55437 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract:

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2. BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-236 and 473-708. The polypeptides sequences are designated SEQ ID NO: 237-472 and 709-944. The nucleic acids and polypeptides are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-236 and 473-708 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-236 and 473-708. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-236 and 473-708 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO:1-236 and 473-708 . The sequence information can be a segment of any one of SEQ ID NO:1-236 and 473-708 that uniquely identifies or represents the sequence information of SEQ ID NO:1-236 and 473-708.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing

full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO:1-236 and 473-708 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO:1-236 and 473-708 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO:1-236 and 473-708; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO:1-236 and 473-708; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO:1-236 and 473-708. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO:1-236 and 473-708; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in SEQ ID NO:237 – 472 or 709-944; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO:1-236 and 473-708; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can

effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady

```
Phe Tyr Tyr Thr Glu Val Gln Leu Lys Glu Glu Ser Ala Ala Ala Ala
                325                 330                 335

gct gct gct gcc gca ggc acc cca gtc cct ggg act ccc acc tcc gag     1174
Ala Ala Ala Ala Ala Gly Thr Pro Val Pro Gly Thr Pro Thr Ser Glu
            340                 345                 350

cca gct ccc acc ccc agc atg act ggc ctg cct ctg tct gct ctt cca     1222
Pro Ala Pro Thr Pro Ser Met Thr Gly Leu Pro Leu Ser Ala Leu Pro
        355                 360                 365

cca cct ctg cac aaa gcc cag tcc tcc ggc cca gaa cat cct ggc ccg     1270
Pro Pro Leu His Lys Ala Gln Ser Ser Gly Pro Glu His Pro Gly Pro
    370                 375                 380

gag tcc tcc ctg ccc tca ggg gct ctc agc aag tca gct cct ggg tcc     1318
Glu Ser Ser Leu Pro Ser Gly Ala Leu Ser Lys Ser Ala Pro Gly Ser
385                 390                 395                 400

ttc tgg cac att cag gca gat cat gca tac cag gct ctg cca tcc ttc     1366
Phe Trp His Ile Gln Ala Asp His Ala Tyr Gln Ala Leu Pro Ser Phe
                405                 410                 415

cag atc cca gtc tca cca cac atc tac acc agt gtc agc tgg gct gct     1414
Gln Ile Pro Val Ser Pro His Ile Tyr Thr Ser Val Ser Trp Ala Ala
            420                 425                 430

gcc ccc tcc gcc gcc tgc tct ctc tct ccg gtc cgg agc cgg tcg cta     1462
Ala Pro Ser Ala Ala Cys Ser Leu Ser Pro Val Arg Ser Arg Ser Leu
        435                 440                 445

agc ttc agc gag ccc cag cag cca gca cct gcg atg aaa tct cat ctg     1510
Ser Phe Ser Glu Pro Gln Gln Pro Ala Pro Ala Met Lys Ser His Leu
    450                 455                 460

atc gtc act tct cca ccc cgg gcc cag agt ggt gcc agg aaa gcc cga     1558
Ile Val Thr Ser Pro Pro Arg Ala Gln Ser Gly Ala Arg Lys Ala Arg
465                 470                 475                 480

ggg gag gct aag aag tgc cgc aag gtg tat ggc atc gag cac cgg gac     1606
Gly Glu Ala Lys Lys Cys Arg Lys Val Tyr Gly Ile Glu His Arg Asp
                485                 490                 495

cag tgg tgc acg gcg tgc cgg tgg aag aag gcc tgc cag cgc ttt ctg     1654
Gln Trp Cys Thr Ala Cys Arg Trp Lys Lys Ala Cys Gln Arg Phe Leu
            500                 505                 510

gac tga gctgtgctgc aggttctact ctgttcctgg ccctgccggc agccactgac     1710
Asp  *
```

<210> 223
<211> 4063
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (158)..(2479)

```
                245                 250                 255

agt gag ccc ttt gtg caa aaa ctc tgg gaa caa tac atg gat gag aag      816
Ser Glu Pro Phe Val Gln Lys Leu Trp Glu Gln Tyr Met Asp Glu Lys
            260                 265                 270

gac gag tac tta cag cag cta aag cag gag ctt ggc ata gaa ctc cat      864
Asp Glu Tyr Leu Gln Gln Leu Lys Gln Glu Leu Gly Ile Glu Leu His
        275                 280                 285

gag gaa gtg act ctg ccc aag ctg cga ggg ggc ctg atg acc atc gac      912
Glu Glu Val Thr Leu Pro Lys Leu Arg Gly Gly Leu Met Thr Ile Asp
    290                 295                 300

ccc agc ctg gac aag cag aca gtg aac acc tac atg agc cag gcc ttc      960
Pro Ser Leu Asp Lys Gln Thr Val Asn Thr Tyr Met Ser Gln Ala Phe
305                 310                 315                 320

cag ctc cct gag tcg gaa atg cca gag gag ggt gac gag aag gaa gaa     1008
Gln Leu Pro Glu Ser Glu Met Pro Glu Glu Gly Asp Glu Lys Glu Glu
                325                 330                 335

gcc gtg gtg gaa atc ctc cag act gcc ctg gag cgg ctt cag gtg att     1056
Ala Val Val Glu Ile Leu Gln Thr Ala Leu Glu Arg Leu Gln Val Ile
            340                 345                 350

gac atc agg cgt gtg gga cct cga gag cca gag cct gca agc tag         1101
Asp Ile Arg Arg Val Gly Pro Arg Glu Pro Glu Pro Ala Ser  *
        355                 360                 365
```

<210> 222
<211> 1710
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (119)..(1660)

<400> 222

```
taagcttgcg gccgctgcag cccggagctg agctagccgt ccgagccgag ccgtccgagc      60

cggggaagcc gggcgcgtgc tgccgctcgt ggcggccgag gagaggagag gcagcagc       118
atg gcg agt gtc ctg tcc cga cgc ctt gga aag cgg tcc ctc ctg gga       166
Met Ala Ser Val Leu Ser Arg Arg Leu Gly Lys Arg Ser Leu Leu Gly
  1               5                  10                  15

gcc cgg gtg ttg gga ccc agt gcc tcg gag ggg ccc tcg gct gcc cca       214
Ala Arg Val Leu Gly Pro Ser Ala Ser Glu Gly Pro Ser Ala Ala Pro
            20                  25                  30

ccc tcg gag cca ctg cta gaa ggg gcc gct ccc cag cct ttc acc acc       262
Pro Ser Glu Pro Leu Leu Glu Gly Ala Ala Pro Gln Pro Phe Thr Thr
        35                  40                  45

tct gat gac acc ccc tgc cag gag cag ccc aag gaa gtc ctt aag gct       310
Ser Asp Asp Thr Pro Cys Gln Glu Gln Pro Lys Glu Val Leu Lys Ala
    50                  55                  60

ccc agc acc tcg ggc ctt cag cag gtg gcc ttt cag cct ggg cag aag       358
```

```
Pro Ser Thr Ser Gly Leu Gln Gln Val Ala Phe Gln Pro Gly Gln Lys
 65                  70                  75                  80

gtt tat gtg tgg tac ggg ggt caa gag tgc aca gga ctg gtg gag cag      406
Val Tyr Val Trp Tyr Gly Gly Gln Glu Cys Thr Gly Leu Val Glu Gln
                 85                  90                  95

cac agc tgg atg gag ggt cag gtg acc gtc tgg ctg ctg gag cag aag      454
His Ser Trp Met Glu Gly Gln Val Thr Val Trp Leu Leu Glu Gln Lys
            100                 105                 110

ctg cag gtc tgc tgc agg gtg gag gag gtg tgg ctg gca gag ctg cag      502
Leu Gln Val Cys Cys Arg Val Glu Glu Val Trp Leu Ala Glu Leu Gln
        115                 120                 125

ggc ccc tgt ccc cag gca cca ccc ctg gag ccc gga gcc cag gcc ctg      550
Gly Pro Cys Pro Gln Ala Pro Pro Leu Glu Pro Gly Ala Gln Ala Leu
    130                 135                 140

gcc tac agg ccc gtc tcc agg aac atc gat gtc cca aag agg aag tcg      598
Ala Tyr Arg Pro Val Ser Arg Asn Ile Asp Val Pro Lys Arg Lys Ser
145                 150                 155                 160

gac gca gtg gaa atg gat gag atg atg gcg gcc atg gtg ctg acg tcc      646
Asp Ala Val Glu Met Asp Glu Met Met Ala Ala Met Val Leu Thr Ser
                165                 170                 175

ctg tcc tgc agc cct gtt gta cag agt cct ccc ggg acc gag gcc aac      694
Leu Ser Cys Ser Pro Val Val Gln Ser Pro Pro Gly Thr Glu Ala Asn
            180                 185                 190

ttc tct gct tcc cgt gcg gcc tgc gac cca tgg aag gag agt ggt gac      742
Phe Ser Ala Ser Arg Ala Ala Cys Asp Pro Trp Lys Glu Ser Gly Asp
        195                 200                 205

atc tcg gac agc ggc agc agc act acc agc ggt cac tgg agt ggg agc      790
Ile Ser Asp Ser Gly Ser Ser Thr Thr Ser Gly His Trp Ser Gly Ser
    210                 215                 220

agt ggt gtc tcc acc ccc tcg ccc ccc cac ccc cag gcc agc ccc aag      838
Ser Gly Val Ser Thr Pro Ser Pro Pro His Pro Gln Ala Ser Pro Lys
225                 230         .       235                 240

tat ttg ggg gat gct ttt ggt tct ccc caa act gat cat ggc ttt gag      886
Tyr Leu Gly Asp Ala Phe Gly Ser Pro Gln Thr Asp His Gly Phe Glu
                245                 250                 255

acc gat cct gac cct ttc ctg ctg gac gaa cca gct cca cga aaa aga      934
Thr Asp Pro Asp Pro Phe Leu Leu Asp Glu Pro Ala Pro Arg Lys Arg
            260                 265                 270

aag aac tct gtg aag gtg atg tac aag tgc ctg tgg cca aac tgt ggc      982
Lys Asn Ser Val Lys Val Met Tyr Lys Cys Leu Trp Pro Asn Cys Gly
        275                 280                 285

aaa gtt ctg cgc tcc att gtg ggc atc aaa cga cac gtc aaa gcc ctc     1030
Lys Val Leu Arg Ser Ile Val Gly Ile Lys Arg His Val Lys Ala Leu
    290                 295                 300

cat ctg ggg gac aca gtg gac tct gat cag ttc aag cgg gag gag gat     1078
His Leu Gly Asp Thr Val Asp Ser Asp Gln Phe Lys Arg Glu Glu Asp
305                 310                 315                 320

ttc tac tac aca gag gtg cag ctg aag gag gaa tct gct gct gct gct     1126
```